(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 663 427 B1**

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention of the grant of the patent:
14.11.2018 Bulletin 2018/46

(21) Application number: **04783494.0**

(22) Date of filing: **08.09.2004**

(51) Int Cl.:
A63F 13/213 *(2014.01)* ***A63F 13/5258*** *(2014.01)*

(86) International application number:
PCT/US2004/029264

(87) International publication number:
WO 2005/028055 (31.03.2005 Gazette 2005/13)

(54) METHOD AND APPARATUS FOR ADJUSTING A VIEW OF A SCENE BEING DISPLAYED ACCORDING TO TRACKED HEAD MOTION

VERFAHREN UND VORRICHTUNG ZUR EINSTELLUNG DER ANSICHT EINER DARGESTELLTEN SZENE GEMÄSS VERFOLGTER KOPFBEWEGUNGEN

PROCEDE ET APPAREIL PERMETTANT DE REGLER LA VUE D'UNE SCENE EN COURS D'AFFICHAGE CONFORMEMENT AU MOUVEMENT D'UNE TETE FAISANT L'OBJET D'UNE POURSUITE

(84) Designated Contracting States:
AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HU IE IT LI LU MC NL PL PT RO SE SI SK TR

(30) Priority: **15.09.2003 US 663236**

(43) Date of publication of application:
07.06.2006 Bulletin 2006/23

(73) Proprietor: **Sony Interactive Entertainment Inc.**
Tokyo 108-8270 (JP)

(72) Inventors:
- **LARSEN, Eric, J**
  Foster City, CA 94404-2175 (US)
- **DESHPANDE, Hrishikesh, R**
  Foster City, CA 94404-2175 (US)
- **MARKS, Richard, L.**
  Foster City, CA 94404-2175 (US)

(74) Representative: **D Young & Co LLP**
120 Holborn
London EC1N 2DY (GB)

(56) References cited:
EP-A- 0 869 458
EP-A- 1 180 384
EP-A- 1 279 425
EP-A2- 0 474 307
EP-A2- 1 279 425
US-A- 6 100 895
US-A1- 2001 019 620
US-B1- 6 304 267

EP 1 663 427 B1

Printed by Jouve, 75001 PARIS (FR)

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] This invention relates generally to video processing, and more particularly to an interface that enables controlling a virtual camera through a user's head motion in order to adjust the view being presented during an interactive entertainment application.

2. Description of the Related Art

[0002] The interactive entertainment industry strives to allow users a realistic as possible experience when playing an interactive video game. Currently, the scene views presented on screen during execution of the interactive application do not allow for the definition of a scene view according to actual tracked movement where the movement is captured without the use of markers. The requirement for a user to wear the sometimes awkward markers is a nuisance that has prevented the wide scale acceptance of the applications associated with the markers.

[0003] One attempt to provide a realistic experience is to provide a canned response to a detected movement. That is, a user may be monitored and if the user ducks or jumps a corresponding character of the application ducks or jumps. However, there is no correlation with the user's movement to the scene view being presented on display screen viewed by the user. Thus, in order to change a scene view being presented, the user is left with manipulating a joy stick to change the scene view. Moreover, a user is required to remember a number of abstract commands in order to access the various scene movement capabilities. For example, in order to peer around a corner within a scene, the user may have to key a button sequence in combination with manipulation of the joy stick in order to achieve the desired functionality. As can be appreciated, this manipulation is wholly unrelated to the physical movement, i.e., peering around a corner, tying to be emulated.

[0004] In view of the foregoing, there is a need for providing a method and apparatus configured to tie the actual movement of a user to modify a scene view being presented, without having the user wear markers, during an execution of an interactive entertainment application.

SUMMARY OF THE INVENTION

[0005] Previously proposed arrangements are disclosed by EP 1 279 425 A2, US 2001/019620 A1 and EP 0 474 307 A2.

[0006] Broadly speaking, the present invention fills these needs by providing a method and apparatus that tracks head motion of a user without markers in order to adjust a view-frustum associated with a scene being displayed. It should be appreciated that the present invention can be implemented in numerous ways, including as a method, a system, computer readable medium or a device. Several inventive embodiments of the present invention are described below.

[0007] The invention is defined by the appended claims.

[0008] Other aspects and advantages of the invention will become apparent from the following detailed description, taken in conjunction with the accompanying drawings, illustrating by way of example the principles of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The invention, together with further advantages thereof, may best be understood by reference to the following description taken in conjunction with the accompanying drawings.

Figure 1 is a simplified schematic diagram illustrating a view-frustum.

Figure 2 is a simplified schematic diagram illustrating a virtual space viewpoint which is capable of being set by an application developer in accordance with one embodiment of the invention.

Figure 3 is a simplified schematic diagram illustrating a top view of a world space configuration where a user's relative position within a three-dimensional cube is used to effect a scene being presented during an interactive entertainment application in accordance with one embodiment of the invention.

Figure 4 is a simplified schematic diagram illustrating alternative facial orientations generated for a system configured to adjust a point of view being displayed according to a user's head movement in accordance with one embodiment of the invention.

Figure 5 is a simplified schematic diagram illustrating the generation of a template and the corresponding matching of the template to a region within a frame of video data in accordance with one embodiment of the invention.

Figures 6A and Figure 6B are simplified schematic diagrams illustrating a change in a view-frustum according to a change in a location of a control object relative to a view port in accordance with one embodiment of the invention

Figure 6C is a simplified schematic diagram illustrating the translation of a view frustum with a control objects motion, thereby providing a parallax effect in accordance with one embodiment of the invention.

Figures 7A and 7B illustrate simplified schematic diagrams comparing virtual world views with real world views in accordance with one embodiment of the invention.

Figure 8 is a simplified schematic diagram illustrating view-frustums configured to maintain an object location constant within a view port in accordance with

one embodiment of the invention.

Figure 9 is a simplified schematic diagram illustrating a view port rotation scheme where an object is viewed from different angles in accordance with one embodiment of the invention.

Figure 10 is a simplified schematic diagram illustrating a scheme where a user's head stays in a fixed location but a view-frustum is rotated according to how the user's head moves I accordance with one embodiment of the invention.

Figure 11 is a simplified schematic diagram illustrating the system configured to enable interactive user control to define a visible volume being displayed in accordance with one embodiment of the invention.

Figure 12 is a flow chart diagram illustrating method operations for managing a visible volume display through a view port in accordance with one embodiment of the invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0010] An invention is disclosed for adjusting a point of view for a scene being displayed during an interactive entertainment application according to the head movement of a user. In the following description, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be apparent, however, to one skilled in the art that the present invention may be practiced without some or all of these specific details. In other instances, well known process steps have not been described in detail in order not to obscure the present invention.

[0011] The embodiments of the present invention modify a point of view associated with a virtual camera during an interactive entertainment application through the marker-less tracking of a control object. Thus, the visible scene being presented on a display screen is effected to the actual movement of the control object. That is, the control object is tracked and the movement of the control object is translated to modify a view-frustum defining a visible scene presented on a display screen. For illustrative purposes, the embodiments described herein designate the control object as a user's head. Of course, certain features of a user's head may be tracked, e.g., the face or any other suitable facial feature. Accordingly, rather than using a joy stick controller to move a virtual camera that defines the point of view for the visible scene being presented, a change in the coordinates of a user's head, that is being tracked through an image capture device, results in defining a new view point and subsequently displaying the image data associated with the new view point. As mentioned above, the tracking of the control object is performed without markers affixed to the control object. Thus, the user is not required to wear a device for tracking purposes. One skilled in the art will appreciate that the image capture device may be any suitable camera, e.g., a web camera.

[0012] In one embodiment, the physical movement in the real world, that is associated with the control object being tracked, is transformed to a virtual movement of a virtual camera defining a visible scene. The visible scene in the virtual world is then displayed through a virtual window, then rendered onto a rectangular area with screen coordinates, referred to as the view port. As used herein, the view port may any suitable display screen, e.g., a television monitor, a computer monitor, etc. While the embodiments described herein refer to a video game application, the embodiments may be applied to any suitable interactive entertainment application. Accordingly, with respect to a video game application any suitable video game console, such as the "PLAYSTATION 2"® manufactured by Sony Computer Entertainment Inc. may be incorporated with the embodiments described below. However, the embodiments including a video game console may also include any suitable computing device in place of the video game console. For example, with reference to on-line gaming applications, the computing device may be a server.

[0013] In one embodiment, a video camera is set proximate to a graphical display and pointed at the user to detect user movements. In particular, a change in location associated with the user's head or face is detected by the video camera. Each frame of video is processed to locate the position of the user's head in the image, by matching a portion of the video frame with a face template captured from a specific user, or a canonical face template. The face template is initially captured by the user placing his face within a capture region defined on a display screen. Once the user's face is within the capture region the system is signaled so that an image of the user's face may be stored as gray scale image data or some other suitable image data for storage in memory. The virtual viewpoint and view frustum used for displaying the scene are modified to correspond to the user's tracked head or face location during execution of an interactive entertainment application. In addition, distance of the user's head from the camera may be determined from the scale of their face/head features in the video image. The mapping from the head location to the virtual view is dependent on the application. For example, a game developer may decide on the factors defining the mapping of the head location to the virtual view.

[0014] Figure 1 is a simplified schematic diagram illustrating a view-frustum. As is generally known, the view-frustum is used to define viewable objects for presentation. Thus, from viewpoint 100 a pyramid is defined. View-frustum 106 is bounded by the four sides of the pyramid having an apex at viewpoint 100. View-frustum 106 may be thought of as a truncated pyramid where near plane 102 clips the pyramid defined from viewpoint 100 at a front end, i.e., closer to the viewpoint. Far plane 104 clips the pyramid at a far end. Thus, view-frustum 106 defines a truncated pyramid volume, wherein the visible volume for display through a view port is defined within the truncated pyramid volume of view frustum 106. One skilled

in the art will appreciate that the view-frustum enables objects defined in three-dimensional space to be culled in order to present the visible objects on a two-dimensional screen. Consequently, plane 102 may be considered a virtual display screen, i.e., a view port, in which objects defined within view-frustum 106 are presented.

[0015] Figure 2 is a simplified schematic diagram illustrating a virtual space viewpoint which is capable of being set by an application developer in accordance with one embodiment of the invention. Viewpoint 110 is associated with a particular distance to virtual window 108. As a result of that distance, the scene being presented through virtual window 108 may be modified. In other words, distance 111 may be used to define a scale of a scene being displayed in virtual window 108. As can be seen, viewpoint 110 may be moved closer to or farther from virtual window 108 in order to manipulate the scene being presented. With reference to video game applications, distance 111 is set by a game developer or programmer as desired. Thus, distance 111 may be manipulated to provide the effect of being right up against virtual window 108, a significant distance away from virtual window 108, or any distance in between. For example, with respect to a video game application, a character running may be associated with a distance being relatively close to virtual window 108 in order to see the ground directly in front of the character. Alternatively, an application displaying a view from an airplane in flight, may be associated with a large distance to provide the effect of a global view.

[0016] Figure 3 is a simplified schematic diagram illustrating a top view of a world space configuration where a user's relative position within a three-dimensional cube is used to effect a scene being presented during an interactive entertainment application in accordance with one embodiment of the invention. Here, image capture device 116 is configured to track control object 112, e.g., a user's head or face region, within a capture zone defined by three-dimensional cube 114. As will be explained in more detail below, image capture device 116 is in communication with a computing device controlling the image data, i.e., the scenes, being presented on display screen 118. Thus, as control object 112 moves within the capture zone the change in location captured by image capture device 116 is translated in order to effect a corresponding change to a view-frustum defining the visible scene being presented on display screen 118. For example, the pyramid discussed with reference to Figure 1 having an apex associated with control object 112 defines a view-frustum. As will be described in more detail below, the movement of control object 112 causes a scene view being presented to change relative to the movement of the control object. In one embodiment, image capture device 116 is configured to zoom-in on various quadrants of three-dimensional cube 114 in order to locate where head 112 is relative to the three-dimensional volume. It should be appreciated that image capture device 116 may be any suitable camera capable of tracking a user's head within a capture zone. It should be further appreciated that the tracking is performed through a marker-less scheme.

[0017] Still referring to Figure 3, in one embodiment, image capture device 116 is a depth camera. For example, the depth camera discussed in US Application Serial Number 10/365,120 entitled "Method and Apparatus for Real Time Motion Capture," is an exemplary depth camera capable of determining a distance of control object 112 to display screen 118. Depth cameras offered by 3DV STSTEMS or CANESTA may be employed as image capture device 116 for the embodiments described herein. Thus, the view being displayed on display screen 118 may be modified in response to movement within capture zone 114 of control object 112. Additionally, the scale associated with the image data being displayed may be manipulated according to a change in the distance of the user's head to the display screen. In another embodiment, an image capture device without depth capability may be used, where control object 112 is a user's head or facial region, and where the size of a face associated with the user's head is compared in successive video frames in order to translate a change in the distance from the user's head to the display screen. For example, a change in size of the user's head, or another suitable control object, within a range of about 15% in the successive video frames may be used to manipulate the scale of a scene being presented on display screen 118. One skilled in the art will appreciate that the use of an image capture device 116 where the image capture device does not have depth capability will require a more powerful processor as compared to the use of a depth camera.

[0018] Figure 4 is a simplified schematic diagram illustrating alternative facial orientations generated for a system configured to adjust a point of view being displayed according to a user's head movement in accordance with one embodiment of the invention. Here, image 120a is initially captured as the template of a user's head. Upon the capture of image 120a, associated images 120b and 120c are generated where the facial orientation is rotated relative to axis 122. That is, image 120b is created by tilting the face of image 120a in one direction, while image 120c is created by tilting the face of image 120a in a different direction, thereby identifying additional three dimensional positions of the head. One skilled in the art will appreciate that numerous other templates may be generated where the orientation or size of the face is modified from the original template. Additionally, any suitable orientation change or size change from the original captured image may be generated. In one embodiment, a degree of change in the orientation or size is determined for use in modifying a scene view.

[0019] Figure 5 is a simplified schematic diagram illustrating the generation of a template and the corresponding matching of the template to a region within a frame of video data in accordance with one embodiment of the invention. Here, template 124a is created upon initialization as described above. In one embodiment, template

124a is a 12x16 pixel size region. Region 130 represents a frame of video data. Within the frame of video data search region 126 is defined. In one embodiment, a size associated with search region 126 is determined by how far, in terms of pixels, a user can move between frames. For example, if a user moves eight pixels in between frames, the size of search region 126 is configured to accommodate this movement so that the movement may be captured. Thus, in one embodiment, each frame of video data is searched within the search region in order to locate a match between the template defined in search region 126 and stored template 124a, thereby enabling computation of a change in movement of a user. In another embodiment, a match is found for template 124a, which is stored in memory as described above, and a corresponding region 124b within search region 126 through a sum of absolute differences scheme also referred to as an L1 norm calculation. That is, values associated with each pixel of template 124a are subtracted from corresponding pixels within an area defined in search region 126 in order to locate an area within the search region that generates the lowest score when compared to template 124a. For example, corresponding pixel values from template 124a and region 124b are subtracted from each other. The absolute value of each of the differences is then taken and summed in order to obtain a score associated with the comparison of template 124a and region 124b. The corresponding region within search area 126 having the lowest score when compared to template 124a is the most likely candidate for a match. In one embodiment, a threshold score must be obtained in order for the region within search area 126 to be considered a match to template 124a. In one embodiment, if no match is found, then the location of the control object defaults to a location determined within a previous frame. One skilled in the art will appreciate that the comparison through the sum of absolute differences is provided for exemplary purposes only and is not meant to be limiting. That is, other suitable techniques such as taking the square of the differences may also be used in order to calculate the score. In essence, any technique which generates a positive value from each of the differences may be used to calculate the score.

[0020] Still referring to Figure 5, in one embodiment, the image data to determine a match are gray scale luminance values associated with each of the pixels of the corresponding image data. It will be apparent to one skilled in the art that other suitable values associated with the pixels may also be used for the calculation to determine a match to template 124a. It should be appreciated that search region 126 may be set to a new default location within display region 130 during the execution of the interactive entertainment application. In addition, the image data used for the template may be dynamic in order to enhance the tracking of the user's facial features. Thus, when tracking the facial region of a user's head, should the user turn his face from the capture device, the image data captured when the facial region was lost may be tracked in substitution of the initial facial region.

[0021] Figures 6A and Figure 6B are simplified schematic diagrams illustrating a change in a view-frustum according to a change in a location of a control object relative to a view port in accordance with one embodiment of the invention. In Figure 6A, the user's head 134a is tracked at an initial position, thereby defining a view-frustum defining visible volume 136 which is behind view port 132 and between side boundaries 136a and 136b. Should the user's head move closer to view port 132, as illustrated by location 134b of the user's head, the associated view-frustum is modified. That is, view-frustum 138, which is defined behind view port 132 and between side boundaries 138a and 138b, provides a wider angle of view behind relative to view-frustum 136. It should be appreciated that this effect may be analogized to looking out a window. That is, the farther a person is from the actual window the view angle will be more limited.

[0022] Figure 6B illustrates the movement of a view-frustum from side to side in an asymmetric manner in accordance with one embodiment of the invention. Here, the user's head is initially in location 133a, thereby defining view-frustum 142 behind view port 132 and between side boundaries 142a and 142b. As the location of the user's head moves to location 133b, the boundaries of view-frustum 140 are modified as compared to view-frustum 142. That is, the user in location 134b has a wider angle of sight through the right-hand side of view port 132 as defined by side boundary 140b. However, a more limited angle of sight through the left-hand side of view port 132 is associated with view frustum 140 through side boundary 140a. It should be appreciated that with reference to Figure 6B, view-frustum 142 defines a symmetrical view-frustum. That is, the line of sight from a user's head at location 133a is normal to a center point 135 of the plane defined by view port 132. However, as the user's location is moved to location 133b the view-frustum is adjusted as described above and becomes asymmetrical. As such, the eye-gaze direction from location 134b is not normal relative to a center of view port 132. In other words, the view plane is no longer perpendicular to the gaze direction, which is atypical for views provided through video games. The display may be considered as a virtual window into a scene, wherein the embodiments described above adjust the view-frustum to show what a user can see through this window as his head moves. It should be appreciated that this provides a parallax effect but also a change in viewing angle.

[0023] Figure 6C is a simplified schematic diagram illustrating the translation of a view frustum with a control objects motion, thereby providing a parallax effect in accordance with one embodiment of the invention. Here, as a user's head, or a facial feature of the user's head, moves from location 135a to 135b, the perpendicular angle of gaze direction to a center point of the corresponding view port is maintained. Thus, it appears as viewpoint 132 is translated along with the change in location of the user's head, which may be referred to as strafing. It

should be appreciated that visible volume captured through the corresponding view frustums of locations 135a and 135b changes as the boundaries of the corresponding view frustums move. In one embodiment, a user's head moving up and down is tracked to cause the game's view frustum to provide a different viewpoint while maintaining a symmetrical view-frustum. One skilled in the art will appreciate that it is important to maintain the virtual camera direction as described in the embodiment of Figure 6c for peering around a corner such as in a stealth or first person shooter game. One skilled in the art will appreciate that a TV screen acting as view port 132 may be viewed much larger than it actually is relative to the entire visual field of the user in order to provide a bigger window for the user so that a user does not feel they are looking through a tiny window.

[0024] Figures 7A and 7B illustrate simplified schematic diagrams comparing virtual world views with real world views in accordance with one embodiment of the invention. In figure 7A, a virtual world view defined by view-frustums originating from locations 144a and 144b through virtual view port 142a. In Figure 7B, real world views are defined by view-frustums associated with location 144a' and 144b' through view port 142b. It should be appreciated that location 144a in the virtual world corresponds to location 144a' in the real world. Likewise, location 144b corresponds to location 144b'. Furthermore, with respect to video game applications or any other interactive entertainment applications, view port 142b may be considered a television screen or any other suitable type of display screen. For Figure 7A, a virtual camera is associated with locations 144a and 144b. In real world configuration of Figure 7B, a tracking device such as a camera tracks movement of the user's head from initial location 144a' to a next location 144b'. This movement is interpreted in the real world as set by code developers for that scene. A physical movement in the real world of Figure 7B is then transformed or mapped into virtual movement in the virtual world of Figure 7A in order to move the virtual camera to define a scene to be displayed on view port 142b in the real world. It should be appreciated that the scale of movement does not necessarily match between the real world and the virtual world. However, the user is provided with the impression of control over the view movement during execution of the interactive entertainment application.

[0025] Figure 8 is a simplified schematic diagram illustrating view-frustums configured to maintain an object location constant within a view port in accordance with one embodiment of the invention. Here, object 150 defines a center of interest point. That is, the view-frustums associated with various locations, such as locations 144a' and 144b' relative to view port 142 center around object 150. Thus, object 150 appears at a constant position in the scene from the different locations 144a' and 144b'. For example, in a game, if there is a statue that is important for some reason, then the configuration described above enables the statue to be maintained at the center point or point of interest of the scene. Therefore, the user's attention is drawn to the statue even though the scene presentation may not be physically correct. As illustrated in Figure 8, in order to maintain the relative position of object 150, the size of view port 142 is adjusted.

[0026] Figure 9 is a simplified schematic diagram illustrating a view port rotation scheme where an object is viewed from different angles in accordance with one embodiment of the invention. Here, the virtual camera orbits around path 152 in order to define the plurality of view-frustums 154-1 to 154-n, which provide views of object 156 at various angles. Thus, this embodiment may be used when looking over a person's shoulder, or as a person moves around path 152 relative to object 156, i.e., directing an orbiting camera in a 3rd person game. Figure 10 is a simplified schematic diagram illustrating a scheme where a user's head stays in a fixed location but a view-frustum is rotated according to how the user's head moves. For example, the user's head may tilt or twist within a location thereby defining different view-frustums. Here, view-frustums 162-1 through 162-n are defined around location 160 which corresponds to a user's head. In this embodiment, a user provided with the capability of looking around a cockpit for flight simulation applications or out of side windows of a vehicle during driving simulation applications.

[0027] Figure 11 is a simplified schematic diagram illustrating the system configured to enable interactive user control to define a visible volume being displayed in accordance with one embodiment of the invention. Here, display device 164 is in communication with computing device 168 which includes a controller 170. Camera 116 is configured to monitor user 172. That is, as user 172 moves, camera 166 tracks a location of the movement of the facial region 174 of the user as described above. A position of virtual camera 176 capturing a scene being presented is adjusted in response to the movement of the facial region, thereby modifying the scene presented through display device 164. For example, camera 116 may be configured to track a user's head located through comparison with a template which is stored in memory of computing device 168. Computing device 168 compares the template to video frame data captured through camera 116 as described with reference to Figure 5. For example, if user 172 moves his head to peer around a corner, virtual camera 176 is adjusted to provide a view on display device 164 providing a scene of what is around the corner.

[0028] Figure 12 is a flow chart diagram illustrating method operations for managing a visible volume display through a view port in accordance with one embodiment of the invention. The method initiates with method operation 180 where a head of the user is identified. For example, the head of the user may be initialized as described above in order for a template to be generated of the head of the user for use as described below. In one embodiment, the initialization of the head of the user cap-

tures a gray scale image of the head of the user and stores that image in memory. The visible volume may be a portion of a view-frustum that defines a scene for presentation as described with reference to Figure 1. The method then advances to operation 182 where a location of the head of the user is tracked relative to a view port. For example, as a user tilts, rotates or moves their head from one location to another, the new location or orientation is tracked relative to a view port. As described above, a view port may be a television screen or any other suitable display screen. Additionally, the movement of the head of the user is captured through a camera that may or may not have depth capturing capability. The method then advances to operation 184 where a view-frustum is translated in accordance with a change in location of the head. Any number of translations of the view-frustum may be used as described above with reference to Figures 7 through 10. Additionally, while a template of a head is used for tracking purposes, one skilled in the art will appreciate that numerous other schemes may be incorporated in place of the template of the head. For example, any suitable marker-less scheme that determines where a user's head may be utilized. In one embodiment, the relative distance of the user's head to a view port is also tracked to adjust a scale associated with scene being presented.

[0029] In summary, the above described embodiments enable the tracking of a user's head in order to move a point of view to correlate to the head movement. The tracking is performed without markers, thereby freeing previous restrictions on a user, especially with reference to virtual reality applications. While exemplary applications have been provided for viewing control applications, it should be appreciated that numerous other suitable applications may use the embodiments described herein. For example, additional specific uses include: directing the view change in a 3D cut-scene, movie or replay; judging distances using the depth-cue provided by head-motion parallax, e.g., what distance to jump in a platformer game; a scary effect of restricting the normal field of view, e.g., showing the small area visible with a flashlight, and requiring the user to move their head to see more; and using head motion as a trigger for events related to the user's view, such as, triggering a warping effect when a user looks over a precipice to indicate vertigo, and triggering a game character's reaction when you look at something. In another embodiment, a sniper mode may be provided where the virtual camera is finely moved according to fine movements of the user's head, similar to peering through the crosshairs of a rifle when aiming the rifle.

[0030] It should be appreciated that the embodiments described herein may also apply to on-line gaming applications. That is, the embodiments described above may occur at a server that sends a video signal to multiple users over a distributed network, such as the Internet, to enable players at remote noisy locations to communicate with each other. It should be further appreciated that the embodiments described herein may be implemented through either a hardware or a software implementation. That is, the functional descriptions discussed above may be synthesized to define a microchip configured to perform the functional tasks for locating and tracking a user's head or facial region and translating the tracked movement to define a scene for presentation.

[0031] With the above embodiments in mind, it should be understood that the invention may employ various computer-implemented operations involving data stored in computer systems. These operations include operations requiring physical manipulation of physical quantities. Usually, though not necessarily, these quantities take the form of electrical or magnetic signals capable of being stored, transferred, combined, compared, and otherwise manipulated. Further, the manipulations performed are often referred to in terms, such as producing, identifying, determining, or comparing.

[0032] The above described invention may be practiced with other computer system configurations including hand-held devices, microprocessor systems, microprocessor-based or programmable consumer electronics, minicomputers, mainframe computers and the like. The invention may also be practiced in distributing computing environments where tasks are performed by remote processing devices that are linked through a communications network.

[0033] The invention can also be embodied as computer readable code on a computer readable medium. The computer readable medium is any data storage device that can store data which can be thereafter read by a computer system. Examples of the computer readable medium include hard drives, network attached storage (NAS), read-only memory, random-access memory, CD-ROMs, CD-Rs, CD-RWs, magnetic tapes, and other optical and non-optical data storage devices. The computer readable medium can also be distributed over a network coupled computer system so that the computer readable code is stored and executed in a distributed fashion.

Claims

1. A method for processing interactive user control with a scene of a video clip, comprising:

identifying (180) a head (112) of a user that is to interact with the scene of the video clip;
tracking (182) the identified head of the user during display of the video clip, the tracking enabling detection of a change in position of the head of the user for each frame of the video clip, the tracking including:

identifying a search region (126) within a portion of a frame of the video clip; and
comparing values within the search region to template values; and

adjusting a view-frustum in accordance with the change in position of the head of the user for each frame of the video clip,

wherein identifying the head of the user includes:

capturing an image (120a) of a face for the head;
generating a template of the face using the captured image; **characterised by**:

generating associated images (120b, 120c) where the facial orientation is tilted relative to the captured image; and
generating other templates using the generated associated images.

2. The method of claim 1, wherein a view-frustum is initially defined by a triangular gaze projection set between outer edges of a virtual window and a virtual position of the head when the virtual position of the head is normal to a center point of the virtual window.

3. The method of claim 2, wherein adjusting the view-frustum moves the virtual position of the head at an angle from normal relative to the center point of the virtual window.

4. The method of claim 3, wherein the virtual position of the head being at the angle from normal relative to the center point of the virtual window changes an angle of the triangular gaze projection, the change in angle of the triangular gaze projection displays a change in viewing angle of the scene provided by the video clip.

5. The method of claim 4, wherein the change in viewing angle of the scene is a result of the detected movement of the head of the user to enable the interaction with the scene.

6. The method of claim 1, wherein the change in position of the head of the user is continuously tracked frame-by-frame of the video clip to enable continual adjustments of the view-frustum so that a viewing angle of the scene is changed in response to the changes in position of the head of the user.

7. The method of claim 1, wherein capturing the template includes:
obtaining a marker-less image of the face for the head of the user.

8. The method of claim 1, wherein a camera is used to enable the capturing.

9. The method of claim 1, wherein the video clip is of a video game.

10. The method of claim 1, wherein the interaction with the scene by tracking movement of the head of the user is independent of user hand-held controls for interacting with the video game.

11. The method of claim 1, wherein the method operation of tracking the identified head of the user during display of the video clip includes:

tracking a facial portion of the head; and
matching gray scale image data associated with the facial portion to image associated with a template of the facial portion.

12. The method of claim 1, wherein the method operation of adjusting a view-frustum in accordance with the change in position of the head of the user includes:

identifying a point of interest of the scene of the video clip; and
modifying the view-frustum so that the point of interest appears at a constant position within the scene when displayed in successive video clips.

13. A computer readable medium having program instructions for processing interactive user control with a scene of a video clip, comprising:

program instructions for identifying (180) a head (112) of a user that is to interact with the scene of the video clip;
program instructions for tracking (182) the identified head of the user in each frame during display of the video clip, the tracking enabling detection of a change in position of the head of the user for each frame of the video clip, the program instructions for tracking including:

program instructions for identifying a search region (126) within a portion of a frame of the video clip; and
program instructions for comparing values within the search region to template values; and
program instructions for adjusting a view-frustum in accordance with the change in position of the head of the user,

wherein the program instructions for identifying the head of the user are for:

capturing an image (120a) of a face for the head;
generating a template of the face using the captured image; **characterised by**:

generating associated images (120b, 120c) where the facial orientation is tilted relative to the captured template; and
generating other templates using the generated associated images.

14. The computer readable medium of claim 13, wherein the program instructions for identifying the head of the user includes:
program instructions for capturing a template of a face for the head.

15. The computer readable medium of claim 13, wherein the change in position of the head of the user is continuously tracked frame-by-frame of the video clip to enable continual adjustments of the view-frustum after each frame of the video clip so that a viewing angle of the scene is changed in response to the changes in position of the head of the user.

16. The computer readable medium of claim 14, wherein the program instructions for capturing the template includes:
program instructions for obtaining a marker-less image of the face for the head of the user.

17. The computer readable medium of claim 13, wherein the program instructions for tracking the identified head of the user during display of the video clip includes:

program instructions for tracking a facial portion of the head; and
program instructions for matching gray scale image data associated with the facial portion to image associated with a template of the facial portion .

18. The computer readable medium of claim 14, wherein the program instructions for adjusting a view-frustum in accordance with the change in position of the head of the user includes:

program instructions for identifying a point of interest of the scene of the video clip; and
program instructions for modifying the view-frustum so that the point of interest appears at a constant position within the scene when displayed in successive video clips.

Patentansprüche

1. Verfahren zur Verarbeitung interaktiver Benutzersteuerung mit einer Szene eines Videoclips, umfassend:

Identifizieren (180) eines Kopfes (112) eines Benutzers, der mit der Szene des Videoclips interagieren soll;
Verfolgen (182) des identifizierten Kopfes des Benutzers während der Anzeige des Videoclips, wobei das Verfolgen eine Erkennung einer Änderung der Position des Kopfes des Benutzers für jeden Frame des Videoclips ermöglicht, wobei das Verfolgen umfasst:

Identifizieren einer Suchregion (126) innerhalb eines Abschnitts eines Frames des Videoclips; und
Vergleichen von Werten innerhalb der Suchregion mit Vorlagenwerten; und
Anpassen eines Ansichtskegelstumpfs gemäß der Änderung der Position des Kopfes des Benutzers für jeden Frame des Videoclips,

wobei das Identifizieren des Kopfes des Benutzers umfasst:

Aufnehmen eines Bildes (120a) eines Gesichts für den Kopf;
Generieren einer Vorlage des Gesichts unter Verwendung des aufgenommenen Bildes;
gekennzeichnet durch:

Generieren von assoziierten Bildern (120b, 120c), auf welchen die Gesichtsorientierung in Bezug auf das aufgenommene Bild geneigt ist; und
Generieren anderer Vorlagen unter Verwendung der generierten assoziierten Bilder.

2. Verfahren nach Anspruch 1, wobei ein Ansichtskegelstumpf anfänglich durch eine trianguläre Blickprojektion definiert wird, die zwischen Außenkanten eines virtuellen Fensters und einer virtuellen Position des Kopfes festgelegt wird, wenn die virtuelle Position des Kopfes normal auf einen Mittelpunkt des virtuellen Fensters ist.

3. Verfahren nach Anspruch 2, wobei das Anpassen des Ansichtskegelstumpfs die virtuelle Position des Kopfes in einem Winkel von der Normalen in Bezug auf den Mittelpunkt des virtuellen Fensters bewegt.

4. Verfahren nach Anspruch 3, wobei die virtuelle Position des Kopfes, die in einem Winkel von der Normalen in Bezug auf den Mittelpunkt des virtuellen Fensters ist, einen Winkel der triangulären Blickprojektion ändert, wobei die Änderung des Winkels der triangulären Blickprojektion eine Änderung des Betrachtungswinkels des Szene anzeigt, die durch den

Videoclip bereitgestellt wird.

5. Verfahren nach Anspruch 4, wobei die Änderung des Betrachtungswinkels der Szene ein Ergebnis der erkannten Bewegung des Kopfes des Benutzers zum Ermöglichen der Interaktion mit der Szene ist.

6. Verfahren nach Anspruch 1, wobei die Änderung der Position des Kopfes des Benutzers Frame für Frame des Videoclips kontinuierlich verfolgt wird, um fortwährende Anpassungen des Ansichtskegelstumpfs zu ermöglichen, damit ein Betrachtungswinkel der Szene in Reaktion auf die Änderungen der Position des Kopfes des Benutzers geändert wird.

7. Verfahren nach Anspruch 1, wobei das Aufnehmen der Vorlage umfasst:
Abrufen eines markierungsfreien Bildes des Gesichts für den Kopf des Benutzers.

8. Verfahren nach Anspruch 1, wobei eine Kamera zum Ermöglichen des Aufnehmens verwendet wird.

9. Verfahren nach Anspruch 1, wobei der Videoclip von einem Videospiel ist.

10. Verfahren nach Anspruch 1, wobei die Interaktion mit der Szene durch Verfolgen von Bewegung des Kopfes des Benutzers von Handbedienung durch den Benutzer zum Interagieren mit dem Videospiel unabhängig ist.

11. Verfahren nach Anspruch 1, wobei der Verfahrensvorgang des Verfolgens des identifizierten Kopfes des Benutzers während der Anzeige des Videoclips umfasst:

Verfolgen eines Gesichtsabschnitts des Kopfes; und
Vergleichen von Graustufenbilddaten, die mit dem Gesichtsabschnitt assoziiert sind, mit einem Bild, das mit einer Vorlage des Gesichtsabschnitts assoziiert ist.

12. Verfahren nach Anspruch 1, wobei der Verfahrensvorgang des Anpassens eines Ansichtskegelstumpfs gemäß der Änderung der Position des Kopfes des Benutzers umfasst:

Identifizieren eines Punktes von Interesse der Szene des Videoclips; und
derartiges Modifizieren des Ansichtskegelstumpfs, dass der Punkt von Interesse in einer konstanten Position innerhalb der Szene erscheint, wenn in nachfolgenden Videoclips angezeigt.

13. Computerlesbares Medium mit Programmanweisungen zum Verarbeiten interaktiver Benutzersteuerung mit einer Szene eines Videoclips, umfassend:

Programmanweisungen zum Identifizieren (180) eines Kopfes (112) eines Benutzers, der mit der Szene des Videoclips interagieren soll;
Programmanweisungen zum Verfolgen (182) des identifizierten Kopfes des Benutzers während der Anzeige des Videoclips, wobei das Verfolgen eine Erkennung einer Änderung der Position des Kopfes des Benutzers für jeden Frame des Videoclips ermöglicht, wobei die Programmanweisungen zum Verfolgen umfassen:

Programmanweisungen zum Identifizieren einer Suchregion (126) innerhalb eines Abschnitts eines Frames des Videoclips; und
Programmanweisungen zum Vergleichen von Werten innerhalb der Suchregion mit Vorlagenwerten; und
Programmanweisungen zum Anpassen eines Ansichtskegelstumpfs gemäß der Änderung der Position des Kopfes des Benutzers,

wobei die Programmanweisungen zum Identifizieren des Kopfes des Benutzers dienen zum:

Aufnehmen eines Bildes (120a) eines Gesichts für den Kopf;
Generieren einer Vorlage des Gesichts unter Verwendung des aufgenommenen Bildes;
gekennzeichnet durch:

Generieren von assoziierten Bildern (120b, 120c), auf welchen die Gesichtsorientierung in Bezug auf das aufgenommene Vorlage geneigt ist; und
Generieren anderer Vorlagen unter Verwendung der generierten assoziierten Bilder.

14. Computerlesbares Medium nach Anspruch 13, wobei die Programmanweisungen zum Identifizieren des Kopfes des Benutzers umfassen:
Programmanweisungen zum Aufnehmen einer Vorlage eines Gesichts für den Kopf.

15. Computerlesbares Medium nach Anspruch 13, wobei die Änderung der Position des Kopfes des Benutzers Frame für Frame des Videoclips kontinuierlich verfolgt wird, um fortwährende Anpassungen des Ansichtskegelstumpfs nach jedem Frame des Videoclips zu ermöglichen, damit ein Betrachtungswinkel der Szene in Reaktion auf die Änderungen der Position des Kopfes des Benutzers geändert wird.

16. Computerlesbares Medium nach Anspruch 14, wobei die Programmanweisungen zum Aufnehmen der Vorlage umfassen:
Programmanweisungen zum Abrufen eines markierungsfreien Bildes des Gesichts für den Kopf des Benutzers.

17. Computerlesbares Medium nach Anspruch 13, wobei die Programmanweisungen zum Verfolgen des identifizierten Kopfes des Benutzers während der Anzeige des Videoclips umfassen:

Programmanweisungen zum Verfolgen eines Gesichtsabschnitts des Kopfes; und
Programmanweisungen zum Vergleichen von Graustufenbilddaten, die mit dem Gesichtsabschnitt assoziiert sind, mit einem Bild, das mit einer Vorlage des Gesichtsabschnitts assoziiert ist.

18. Computerlesbares Medium nach Anspruch 14, wobei die Programmanweisungen zum Anpassen eines Ansichtskegelstumpfs gemäß der Änderung der Position des Kopfes des Benutzers umfassen:

Programmanweisungen zum Identifizieren eines Punktes von Interesse der Szene des Videoclips; und
Programmanweisungen zum derartigen Modifizieren des Ansichtskegelstumpfs, dass der Punkt von Interesse in einer konstanten Position innerhalb der Szene erscheint, wenn in nachfolgenden Videoclips angezeigt.

Revendications

1. Procédé pour le traitement d'une commande interactive d'utilisateur avec une scène d'un extrait vidéo, comprenant :

l'identification (180) d'une tête (112) d'un utilisateur devant interagir avec la scène de l'extrait vidéo ;
la poursuite (182) de la tête identifiée de l'utilisateur lors de l'affichage de l'extrait vidéo, la poursuite permettant la détection d'un changement de position de la tête de l'utilisateur pour chaque trame de l'extrait vidéo, la poursuite comportant :

l'identification d'une région de recherche (126) à l'intérieur d'une partie d'une trame de l'image de l'extrait vidéo ; et
la comparaison de valeurs à l'intérieur de la région de recherche à des valeurs de modèle ; et
l'ajustement d'un tronc de cône de vue en fonction du changement de position de la tête de l'utilisateur pour chaque trame de l'extrait vidéo,
dans lequel l'identification de la tête de l'utilisateur comprend :

la capture d'une image (120a) d'un visage pour la tête ;
la génération d'un modèle de visage en utilisant l'image capturée ;
caractérisé par :

la génération d'images associées (120b, 120c) dans lesquelles l'orientation du visage est inclinée par rapport à l'image capturée ; et
la génération d'autres modèles en utilisant les images associées générées.

2. Procédé selon la revendication 1, dans lequel un champ de vue est initialement défini par une projection triangulaire du regard placée entre des bords extérieurs d'une fenêtre virtuelle et une position virtuelle de la tête lorsque la position virtuelle de la tête est normale à un point central de la fenêtre virtuelle.

3. Procédé selon la revendication 2, dans lequel l'ajustement du tronc de cône de vue déplace la position virtuelle de la tête d'un certain angle par rapport à la normale au point central de la fenêtre virtuelle.

4. Procédé selon la revendication 3, dans lequel la position virtuelle de la tête se situant audit angle par rapport à la normale au point central de la fenêtre virtuelle modifie un angle de la projection triangulaire du regard, le changement d'angle de la projection triangulaire du regard affiche un changement de l'angle de vue de la scène fournie par l'extrait vidéo.

5. Procédé selon la revendication 4, dans lequel le changement d'angle de vue de la scène est un résultat du mouvement détecté de la tête de l'utilisateur pour permettre l'interaction avec la scène.

6. Procédé selon la revendication 1, dans lequel le changement de position de la tête de l'utilisateur est poursuivi en continu trame par trame de l'extrait vidéo pour permettre des ajustements continus du tronc de cône de vue de manière à ce qu'un angle de vue de la scène soit modifié en réponse aux changements de position de la tête de l'utilisateur.

7. Procédé selon la revendication 1, dans lequel la capture du modèle comporte :
l'obtention d'une image sans marqueur du visage pour la tête de l'utilisateur.

8. Procédé selon la revendication 1, dans lequel une caméra est utilisée pour permettre la capture.

9. Procédé selon la revendication 1, dans lequel l'extrait vidéo fait partie d'un jeu vidéo.

10. Procédé selon la revendication 1, dans lequel l'interaction avec la scène par poursuite du mouvement de la tête de l'utilisateur est indépendante de commandes manuelles de l'utilisateur pour interagir avec le jeu vidéo.

11. Procédé selon la revendication 1, dans lequel le procédé de poursuite de la tête identifiée de l'utilisateur pendant l'affichage de l'extrait vidéo comprend :

la poursuite d'une partie visage de la tête ; et
la mise en correspondance de données d'images en niveaux de gris associées à la partie visage avec une image associée à un modèle de la partie visage.

12. Procédé selon la revendication 1, dans lequel l'opération du procédé consistant à ajuster un tronc de cône de vue en fonction du changement de position de la tête de l'utilisateur comprend :

l'identification d'un point d'intérêt de la scène de l'extrait vidéo ; et
la modification du champ de vue pour que le point d'intérêt apparaisse à une position constante à l'intérieur de la scène lorsqu'il est affiché dans des extraits vidéo successifs.

13. Support lisible par ordinateur comportant des instructions de programme permettant le traitement d'une commande interactive d'utilisateur avec une scène d'un extrait vidéo, comprenant :

des instructions de programme permettant l'identification (180) d'une tête (112) d'un utilisateur devant interagir avec la scène de l'extrait vidéo ;
des instructions de programme permettant la poursuite (182) de la tête identifiée de l'utilisateur dans chaque trame pendant l'affichage de l'extrait vidéo, la poursuite permettant la détection d'un changement de position de la tête de l'utilisateur pour chaque trame de l'extrait vidéo, les instructions de programme pour la poursuite comprenant :

des instructions de programme permettant l'identification d'une région de recherche (126) à l'intérieur d'une partie d'une trame de l'extrait vidéo ; et
des instructions de programme permettant la comparaison de valeurs à l'intérieur de la région de recherche à des valeurs de modèle ; et
des instructions de programme pour l'ajustement d'un tronc de cône de vue en fonction du changement de position de la tête de l'utilisateur, dans lequel les instructions de programme permettant l'identification de la tête de l'utilisateur sont destinées à :

la capture d'une image (120a) d'un visage pour la tête ;
la génération d'un modèle de visage à l'aide de l'image capturée ;
caractérisé par :

la génération d'images associées (120b, 120c) dans lesquelles l'orientation du visage est inclinée par rapport au modèle capturé ; et
la génération d'autres modèles en utilisant les images associées générées.

14. Support lisible par ordinateur selon la revendication 13, dans lequel les instructions de programme permettant l'identification de la tête de l'utilisateur comprennent :
des instructions de programme permettant la capture d'un modèle de visage pour la tête.

15. Support lisible par ordinateur selon la revendication 13, dans lequel le changement de position de la tête de l'utilisateur est poursuivi en continu trame par trame de l'extrait vidéo pour permettre des ajustements continus du champ de vue après chaque trame de l'extrait vidéo de manière à ce qu'un angle de vue de la scène soit modifié en réponse aux changements de position de la tête de l'utilisateur.

16. Support lisible par ordinateur selon la revendication 14, dans lequel les instructions de programme permettant la capture du modèle comprennent :
des instructions de programme permettant l'obtention d'une image sans marqueur du visage pour la tête de l'utilisateur.

17. Support lisible par ordinateur selon la revendication 13, dans lequel les instructions de programme permettant la poursuite de la tête identifiée de l'utilisateur pendant l'affichage de l'extrait vidéo comprennent :

des instructions de programme permettant la poursuite d'une partie de visage de la tête ; et
des instructions de programme permettant la mise en correspondance de données d'images en niveaux de gris associées à la partie de visage avec une image associée à un modèle de

la partie de visage.

18. Support lisible par ordinateur selon la revendication 14, dans lequel les instructions de programme permettant l'ajustement d'un tronc de cône de vue en fonction du changement de position de la tête de l'utilisateur comprennent :

des instructions de programme permettant l'identification d'un point d'intérêt de la scène de l'extrait vidéo ; et
des instructions de programme permettant la modification du tronc de cône de vue de manière à ce que le point d'intérêt apparaisse à une position constante dans la scène lorsqu'il est affiché dans des extraits vidéo successifs.

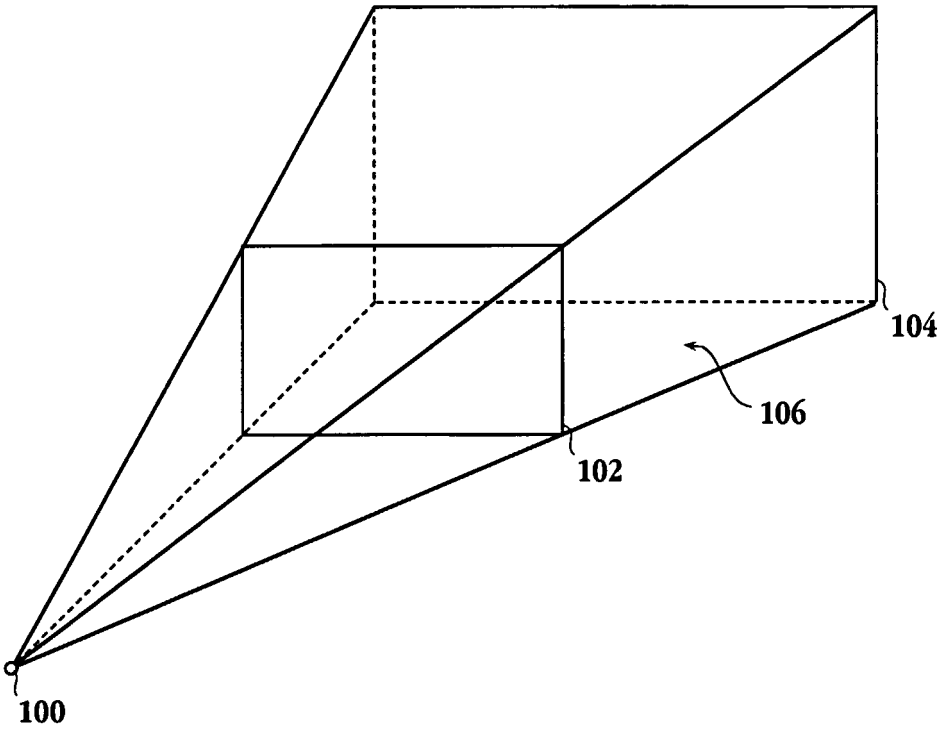

Fig. 1

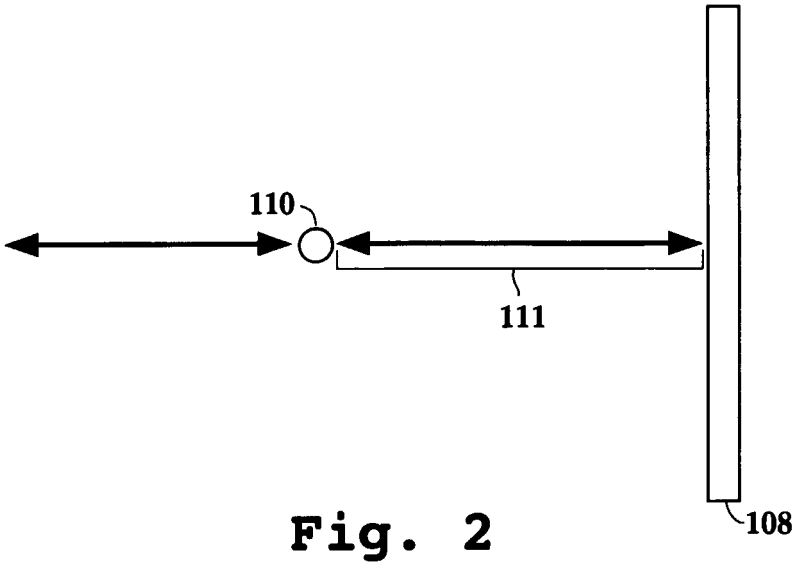

Fig. 2

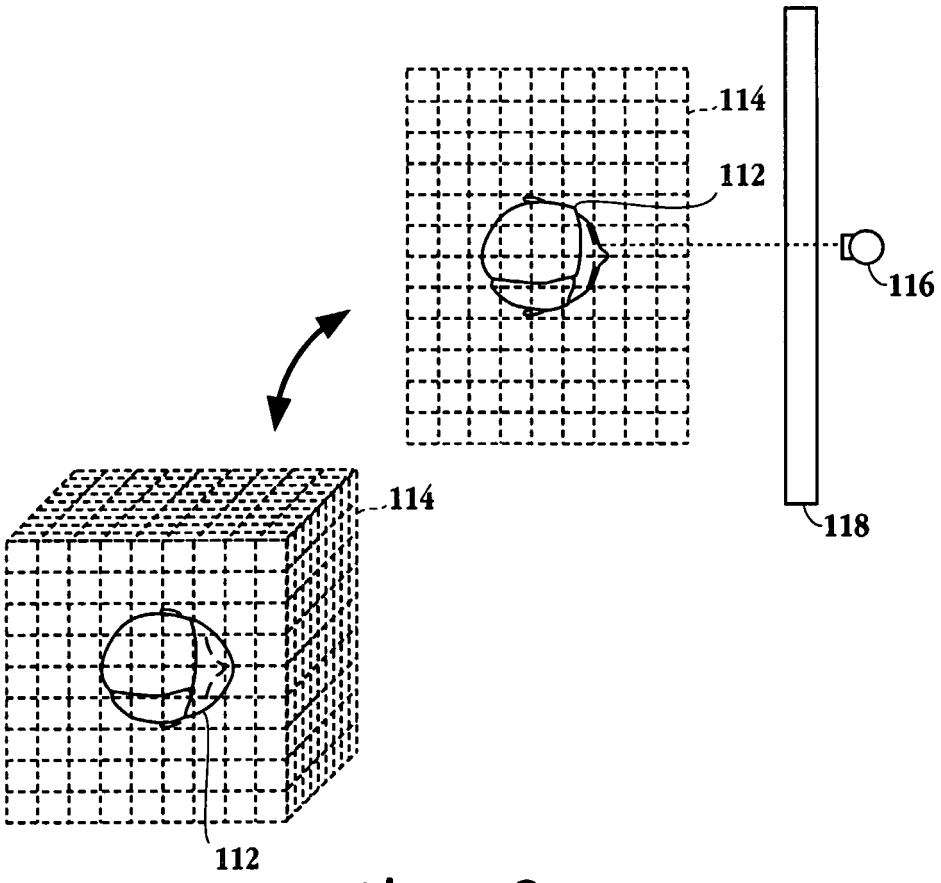

Fig. 3

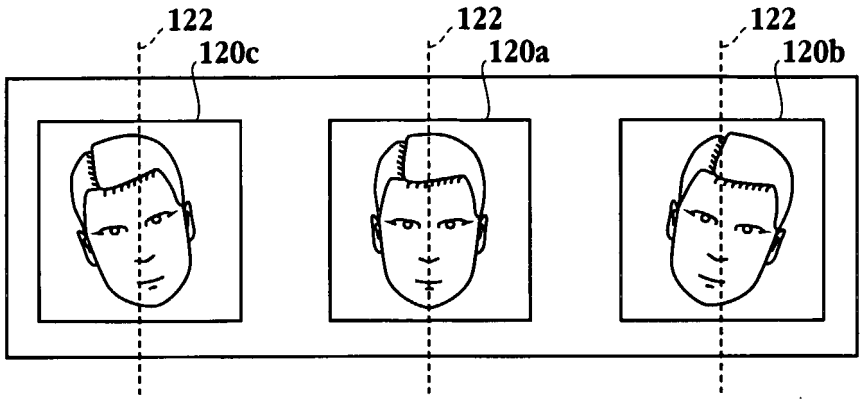

Fig. 4

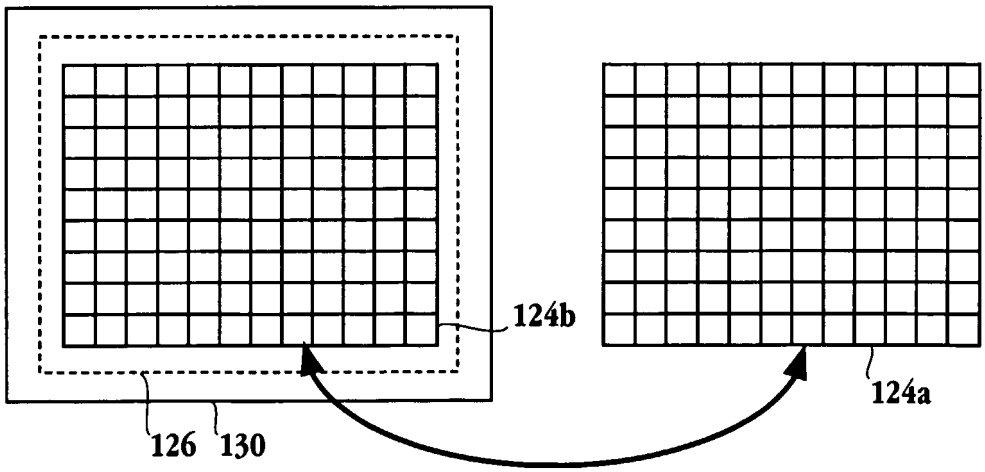

Fig. 5

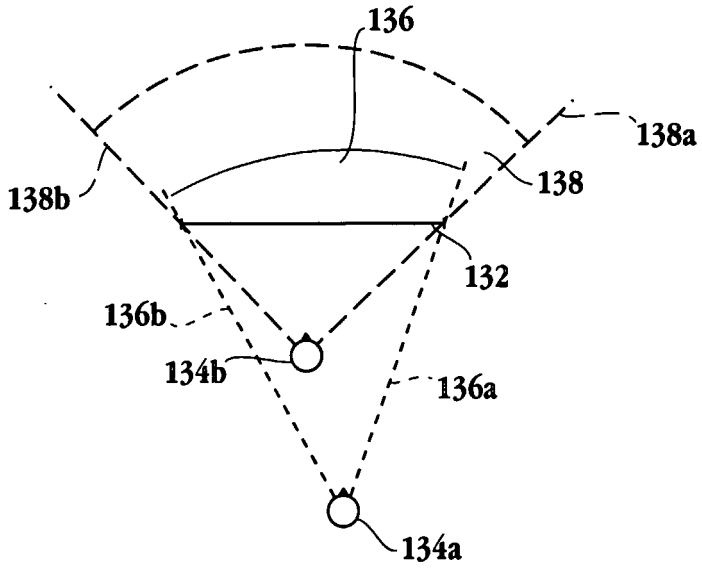
136
138a
138b
138
132
136b
134b
136a
134a

Fig. 6A

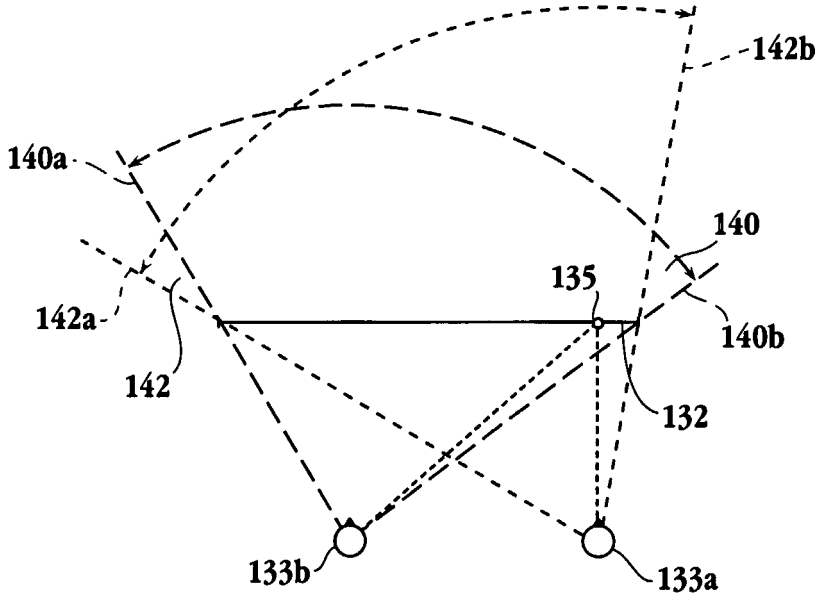
142b
140a
140
135
142a
140b
142
132
133b
133a

Fig. 6B

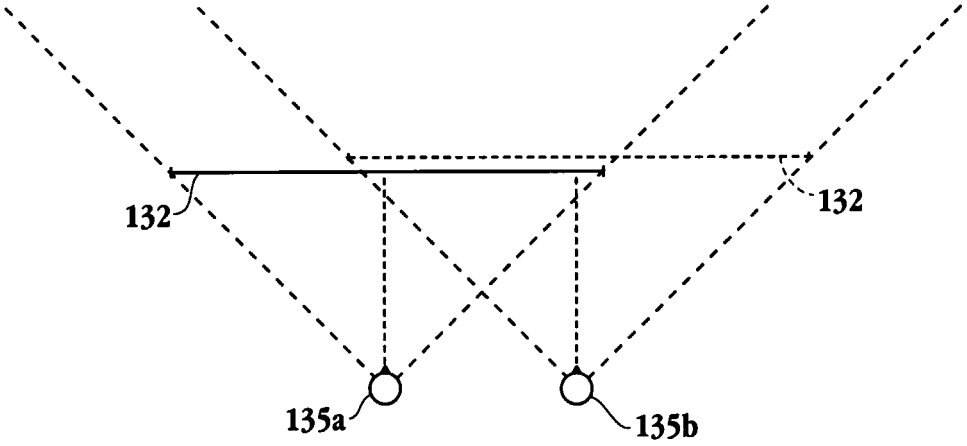

Fig. 6C

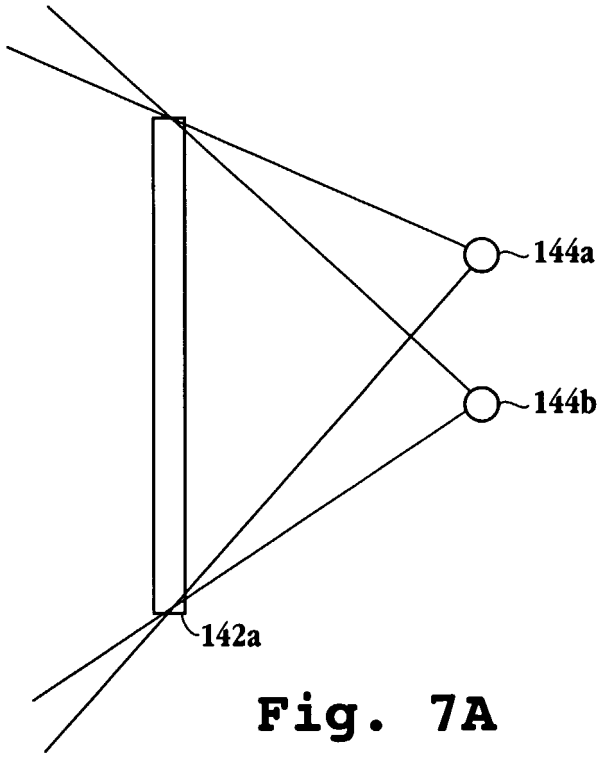

Fig. 7A

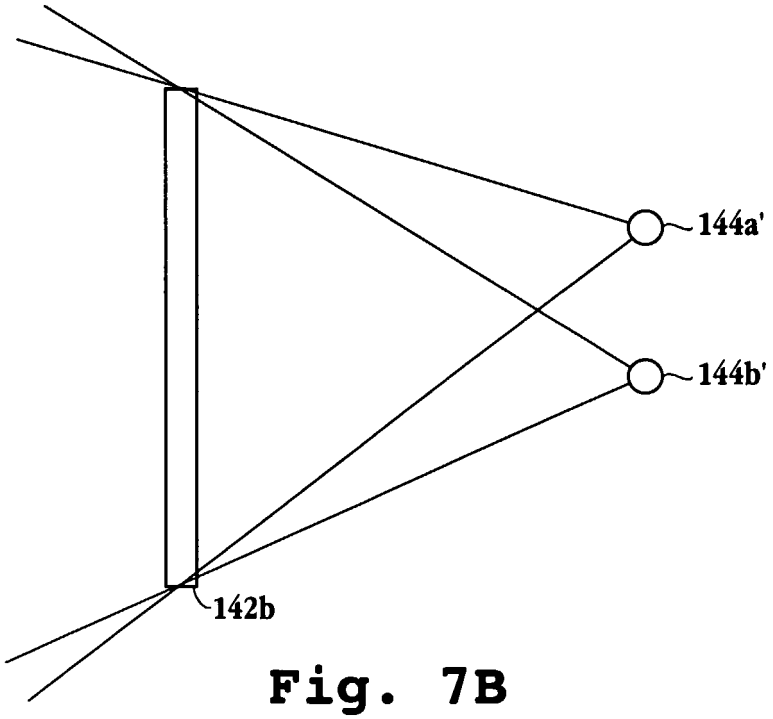

Fig. 7B

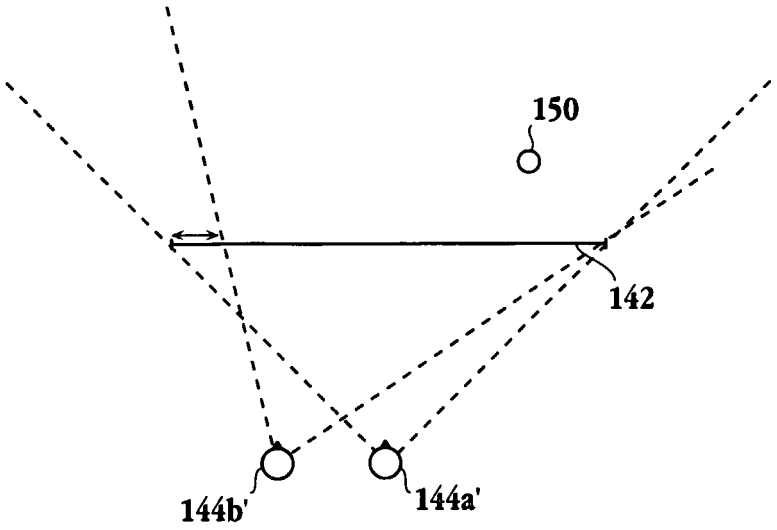

Fig. 8

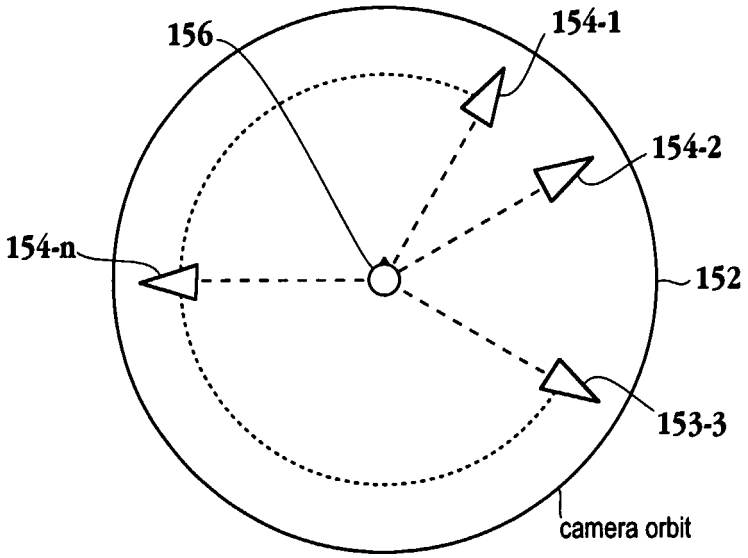

Fig. 9

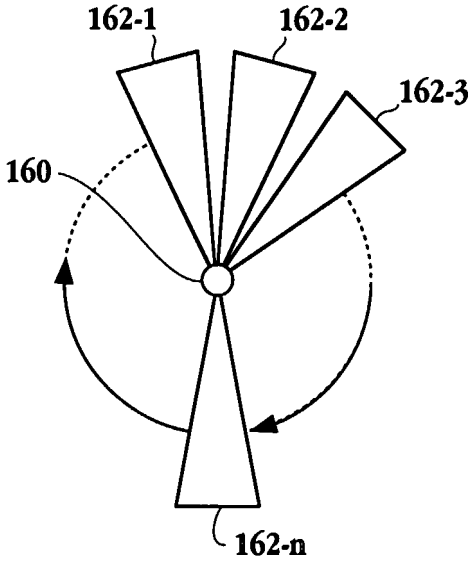

Fig. 10

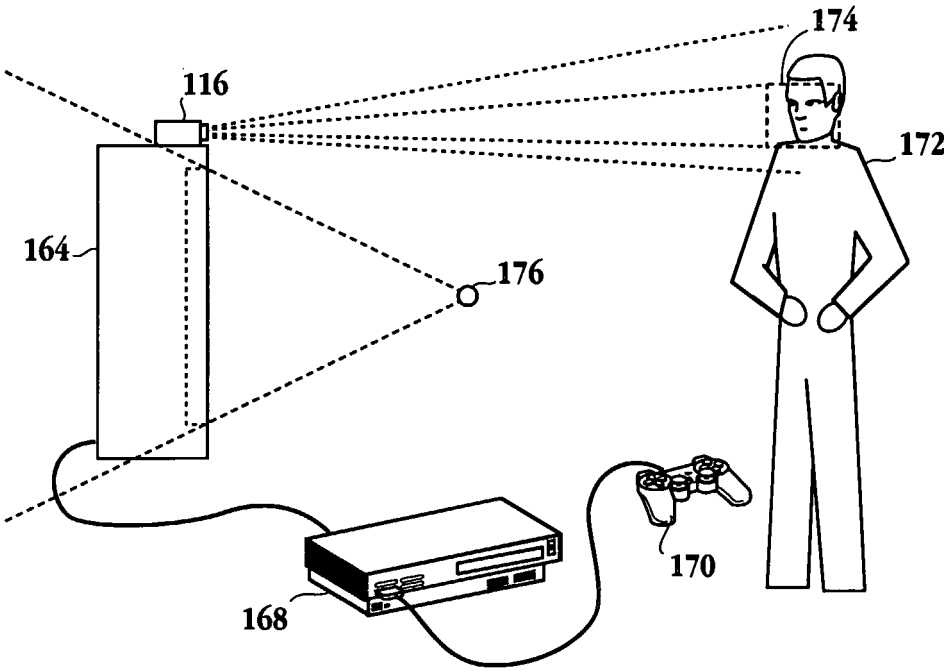

Fig. 11

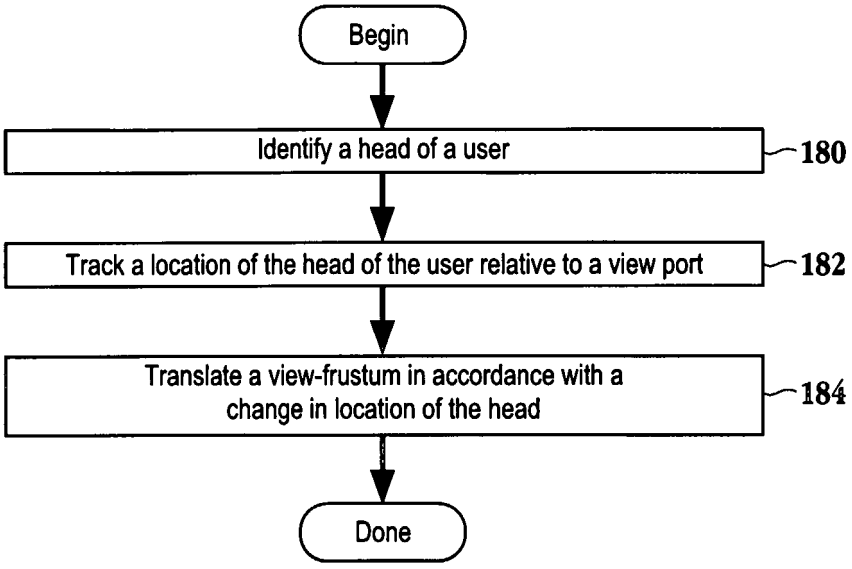
Begin
Identify a head of a user
180
Track a location of the head of the user relative to a view port
182
Translate a view-frustum in accordance with a change in location of the head
184
Done

Fig. 12

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- EP 1279425 A2 **[0005]**
- US 2001019620 A1 **[0005]**
- EP 0474307 A2 **[0005]**
- US 365120 A **[0017]**